DOCUMENT RESUME

ED 064 376 TM 001 621

AUTHOR Woodward, Jane P.; Yeager, John L.
TITLE Evaluation of Programs to Train Educational R&D Personnel.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
PUB DATE Apr 72
NOTE 26p.; Paper presented at the annual meeting of the AERA (Chicago, Ill., April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Concept Teaching; *Educational Programs; *Evaluation Techniques; *Models; Problem Solving; *Program Evaluation; *Research and Instruction Units; Staff Orientation
IDENTIFIERS *CIPP Model

ABSTRACT

A discussion is given of the formative evaluation process as it was conceptualized and conducted to assess a R&D training program under development. Included are a discussion of the evaluation context, the evaluation model, the procedures and instruments developed and implemented and their success, the types of evaluation problems encountered and those conclusions and recommendations which evolved from the project evaluation staff's experience with the formative evaluation effort. The context of the R&D training program evaluation posed a particularly interesting setting and problem for the design and conduct of evaluation. The field of training for the program being developed was relatively undefined in terms of formal conceptualization and existence of research. The model, the CIPP model, comprises four steps: context evaluation, input evaluation, process evaluation, and product evaluation. Evaluation procedures included monitoring techniques, student opinion self-report techniques, achievement assessment techniques, and student attitude assessments. Results include: (1) Too much time was spent on evaluation activities requesting redundant data; (2) Oral feedback was preferable; (3) Time was a problem in the evaluation procedure. Conclusions include: (1) Formative evaluation must be "adaptive" evaluation; (2) The planning of the evaluation design and procedures must take into account any known characteristics of the particular students who comprise the pilot test group; and (3) The CIPP model can serve as a useful base for an overall program evaluation design. (CK)

ED 064376

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EVALUATION OF PROGRAMS TO TRAIN EDUCATIONAL R&D PERSONNEL

Jane P. Woodward and John L. Yeager

Learning Research and Development Center

University of Pittsburgh

TM 001 621

April, 1972

Presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April 6, 1972. The research and development reported herein was performed as part of the R&D Project at the Learning Research and Development Center, University of Pittsburgh.

EVALUATION OF PROGRAMS TO TRAIN EDUCATIONAL R&D PERSONNEL

by

Jane P. Woodward and John L. Yeager

One goal of the University of Pittsburgh R&D Training Project is the development of tested approaches to training program evaluation. The existance of this goal is in recognition of the verity of such statements as the one appearing in the September, 1970 issue of Educational Comment, stating that:

> Each year a total of 4.5 billion dollars is spent by the Federal government on education, and an estimated 40.6 billion dollars by all public schools in the United States. There is an urgent need to develop procedures which will help to assure that dollars for education are being spent wisely. There is an urgent need to develop systems and procedures for improving the evaluation of education. There is an urgent need to develop procedures for evaluating the effectiveness of various instructional programs. (Alkin, 1970)

Changes occuring within the field of educational evaluation itself have sharpened these needs. Attention has been shifting gradually from an emphasis upon the existing classical, experimental "summative" evaluation procedures, where a completed educational product (e.g. set of materials) is evaluated by comparing its effectiveness with that of other treatments, to an emphasis on formative evaluation, which stresses evaluation of the success of an instructional product during its developmental stage, leading to product revision until it is able to attain its stated objectives. This shift has confronted the educational evaluation field with the task of developing and validating

procedures and instruments appropriate to the formative evaluation process. Although several models and many procedures currently exist, in many cases, their feasibility and usefulness in the context of program development remain to be demonstrated.

The purpose of this particular paper is to address the formative evaluation process as it was conceptualized and conducted to assess a R&D training program under development at the University of Pittsburgh Learning Research and Development Center. It includes a discussion of the evaluation context the evaluation model, the procedures and instruments developed and implemented and their success, the types of evaluation problems encountered, and those conclusions and recommendations which evolved from the project evaluation staff's experience with the formative evaluation effort. The majority of the discussion and examples used in this paper will center on Program 3A, the short-term training program on local change which has previously been described by my colleagues.

Evaluation Context

Educational evaluation always occurs within an educational context which imposes certain restrictions and demands upon the type of evaluation procedures that can be utilized. To the extent that this context varies, so must the evaluation being conducted vary. Specific instructional situations and programs possess unique characteristics and evaluators must recognize this uniqueness and be responsive in terms of the kinds of procedures and assessment instruments utilized, the nature of the information collected, and the manner of reporting that information.

The context of the R&D Training Program evaluation posed a particularly interesting setting and problem for the design and conduct of evaluation. A decision was made to initiate the formal training of students prior to completion of program development. This decision was based on the fact that the field of training for the program being developed was relatively undefined in terms of

(a) formal conceptualization of the scope of knowledge and range of skills required of an acknowledged "expert" and (b) the existence of research, writings, or formal training materials. It was therefore anticipated that the initial training group, comprising individuals actively involved in the field, would be able to provide valuable input in terms of their own expertise and experience and therefore contribute to the design and validation of the training objectives and materials. This decision, however, implied a potential compromise of the two major project goals: training of educational personnel, and the development of tested training programs. Concurrent initial development and training meant that a balance had to be negotiated between modifying the training components during program implementation to insure meeting minimal training group needs (the training goal), and trying out and validating the training components for the objectives originally designed (developmental goal) independent of the special needs of the particular group on which the program was being pilot-tested. Therefore a potential incompatability existed between the need to adapt training to student needs and the need to develop the instructional product originally planned. The severity of the conflict depended upon the degree to which the training sample (pilot test group) exhibited the characteristics of the intended target population for the training, the degree to which the training components were individualized and the degree to which the immediate student training needs were viewed as more or less important than program component testing. The evaluation problem posed by the concurrent development and training of students was therefore one of (1) developing procedures sufficiently comprehensive to gather the desired student input for use in developing the program, (2) implementing procedures which would allow immediate feedback for purposes of on-line program modification, and (3) developing an evaluative design and procedure which would be sufficiently flexible to adapt to the major program changes which might occur, particularly if trainee needs were given higher priority.

Another significant aspect of the evaluation context was the role of the evaluation staff which was clearly established as one of providing the materials

development and training staff with information as to program effectiveness. The activities of the evaluation staff were perceived as a service function that was to be responsive to the operational requirements of the program. This characteristic of responsiveness greatly influenced the number and type of measurements initially proposed and the resulting measurements that were ultimately utilized by the program.

The training program was conducted for a six week period with the training period divided into three sessions of 3 weeks, 2 weeks and 1 week, with intervening periods of from 1-3 months. These intervening periods were to be utilized to engage the trainee in active on-the-job training (OJT), involving tryout and practice of the skills learned during the previous training sessions.

Given the project objectives and the constraints that were operating, the evaluation of the project focused on the following:

A. The selection or construction and implementation of an evaluation process based on an existing or newly constructed evaluation model;

B. the development of an evaluation plan, procedures, and instruments appropriate to the program context to include:

   1. process evaluation procedures to monitor program implementation to determine the extent to which the program in operation reflected the original program design, and to provide immediate feedback as to any defects in the training so as to enable on-line modification of the program if needed, and

   2. product evaluation procedures to assess the extent of program effectiveness in attaining its stated goals, providing feedback useful both in determining program effectiveness and in revision of the training programs.

C. the testing and validation of the evaluation procedures proposed and implemented through establishing procedures for evaluating the effectiveness of the program evaluation.

Evaluation Process and Model

The evaluation model which served as the basis of the project evaluation was the CIPP model proposed by Stuffelbeam and best explicated in his most recent publication, Educational Evaluation and Decision-Making (Stuffelbeam, Foley, Gephart, Guba, Hammond, Merriman, and Provus, 1971). This model comprises four steps: (a) context evaluation, which aids planning decisions to determine program objectives by providing the rationale for these decisions; (b) input evaluation, which assists the decision-maker in making structuring decisions regarding determination of program design by identifying and assessing available resources in terms of their potentiality for meeting the objectives identified; (c) process evaluation, which assists decision-making about program operations during program implementation by providing feedback to the decision-makers about defects in procedural design prior to and during program implementation, by providing information for programmed decisions, and by maintaining a record of the program as implemented; and (d) product evaluation, which aids decision-making about program recycling by providing information with regard to attainment of program objectives, both during and at the end of program implementation.

This model, like others, views the primary evaluation function to be the collection and provision of information to assist decision-making concerning program planning, implementation, and revision (Popham, 1971; Stuffelbeam et al., 1971).

Aside from determination of the evaluation design (input evaluation), the main efforts of the program evaluation staff fell within the context of process and product evaluation. Design and implementation of summative evaluation for program certification were postponed until completion of the initial first-year pilot-testing of the programs, and not included in the evaluation design.

Evaluation Procedures

The evaluation design, as based on the CIPP model described above, included the following general categories of procedures:

1. Process evaluation procedures:
   A. Monitoring techniques.
   B. Student opinion self-report techniques, including those to provide:
      (a) Immediate feedback into on-line program modification
      (b) Delayed feedback for subsequent program revision.
   C. Achievement assessment techniques.
2. Product evaluation procedures:
   A. Achievement assessment techniques.
   B. Student attitude assessments.

The particular evaluation context, i.e. the concurrent initial program development and training of students, the varied kinds of objectives, and dictated a comprehensive evaluation design. To guarantee such comprehensiveness, several procedures were prescribed initially to collect each kind of information needed. The specific techniques utilized (with reference to Program 3A only) are described in Appendix A.

RESULTS

Evaluation Procedures

Exhibit 1 presents the evaluation procedures utilized in Training Program 3A. The general procedures used in each session are listed on the left in the temporal order of their administration, while those on the right are those procedures implemented daily during the particular session involved.

Since one intent of the evaluation component was to assess the relevant merit and feasibility of using selected data gathering techniques, the original evaluation design provided several procedures for collecting each kind of data needed. This redundancy was also necessitated by the concurrent implementation of student training and initial program development. Furthermore, by inserting redundancy into the system it was anticipated that partial reliability could be established through the confirmation of information through various data collection techniques. Unit rating sheets, daily logs, section evaluation sheets, the counselor interview, and class feedback sessions, for example, were all assigned to the collection of student opinion data. Student and staff feedback about the evaluation procedures during and upon completion of the initial 3-week session, however, indicated that:

(1) Too much time was spent on evaluation activities requesting redundant data.

(2) The evaluation stressed form completion or comprehensive and lengthly written examination, and there was a decided preference for oral techniques typified by the class feedback sessions.

From the standpoint of the evaluation staff, it became evident that:

(1) The attempt to meet all contingencies that might effect the program resulted in the implementation of a surfeit of procedures to collect the same general kinds of redundant information.

(2) Provision of time during the instructional sessions for completion and collection of the unit ratings and daily logs was a problem, since instructors frequently found it undesirable to

EXHIBIT 1

Program 3A Evaluation Schedule

Session 1 - (Weeks 1-3)	Daily
Introduction (Unit 1) Program and Section 1 pre-tests Section 1 (Units 2-12) Student self-assessment of knowledge (Unit 3) Student self-assessment of interpersonal competencies (Unit 4) Section 1 post-test Section 1 evaluation sheet Section 2 (Units 13-20) Counselor interview Section 2 evaluation sheet	Unit rating sheets Daily logs Evaluator observation and taping of instructional sessions Instructor unit rating sheets Class feedback sessions (as needed)
OJT: Three months	
Daily records Weekly summaries	
Session 2 - (Weeks 4-5)	Daily
Objectives assessment (Units 1-20) Section 3 (Units 21-29) Project evaluation Section 3 evaluation sheet	Taping of instructional sessions Individual feedback Class feedback sessions (as needed) Instructor observation
OJT: One month	
Session 3 - (Week 6)	Daily
Review test Project evaluation 6th week post-test Final evaluation sheet	Taping of instructional sessions Instructor observation Individual feedback Class feedback sessions (as needed)
Follow-Up - Three months later	
Terminal attitude measure	

terminate active discussion for that purpose, and units often ended at lunch hour or extended beyond the appointed afternoon hour. These were, therefore, often filled out and turned in some time subsequent to unit conclusion, limiting their usefulness, reliability, and specificity.

(3) The multiple procedures implemented to collect student opinion data (e.g., daily logs, class feedback sessions, section evaluations, etc.) collected a volume of data which was sufficiently unstructured and difficult to analyze and interpret as to hinder its usefulness and reliability in terms of providing feedback for both immediate and later revision. Procedures to collect more specific, differentiated and controlled data were required which would be more useful for revisions purposes.

(4) Students indicated a definite preference for evaluation techniques which were structured (controlled response) rather than unstructured (free response), and for oral or unobstrusive performance measures in lieu of written evaluation.

(5) Completion of instructor rating sheets was not feasible during program implementation due to severe time constraints on the developmental/instructional staff, resulting in time delays between the end of a lesson and when the form was completed. This technique thus tended to provide only a rather superficial type of data.

(6) Continuation of formal observation of class sessions by the evaluation staff was not feasible in light of the limited staff resources.

(7) The taped interview procedure was less useful than anticipated, due to both the redundancy of the measure and the difficulty of transcribing and accessing the data.

Changes brought about during the first session as a result of student feedback and staff insight included the addition of class feedback sessions, the elimination of the instructor rating sheets and of the Section 2 post-test originally planned for the end of the second section, and the replacement of this posttest with an individual student project design which represented the cumulative set of skills that were provided the student during the first session. Revisions of the evaluation procedures which were manifested in the evaluation design of the second session included:

(1) Elimination of the unit rating sheets and daily logs.

(2) Maintenance of the class feedback session mechanism.

(3) Elimination of the evaluator observation.

(4) Addition of a project evaluation procedure based on the institution of individual student projects during the first session and on the type of objectives (open-ended) dealt with during the second session.

(5) Addition of an objectives assessment procedure to remedy the failure of the student unit ratings from the first session to collect sufficiently specific data about the objectives of the first session's units.

The evaluation design for the second session sharply decreased the number, redundancy, and type of evaluation procedures implemented. In most cases, only one procedure was implemented to collect a given kind of feedback. Project evaluation and instructor observation replaced formal pre- and post-testing, since the session objectives were less susceptible to formal written evaluation. The section evaluation sheet was retained in its original form to gather the generalized student opinion data formerly collected by unit ratings, daily logs, interviews, and section evaluations. The class feedback session was used to provide any immediate feedback necessary for on-line program modification, since it was ascertained that the students could be relied upon to notify the staff when changes were desirable. The instructional taping was retained as the program monitoring technique, and was continued throughout.

The OJT records were discontinued during the second OJT period, for several reasons:

(1) The daily record proved to provide little information which could not be gathered equally well during the sessions themselves.

(2) The second OJT period was brief, only 4 weeks in length.

(3) The amount of time and effort on the part of the student to prepare the materials and the information provided to the staff indicated that this technique had a relatively low efficiency.

Revisions of the program evaluation implemented during the final week's session included:

(1) Administration of the main body of the post-test over the first five-weeks' objectives in the form of a review test administered at the

beginning of the 6th week. This test replaced the section pre- and post-tests, and served to diagnose weaknesses in student mastery of the objectives. Only those objectives failed and the new objectives from the sixth week were then evaluated on the final 6th-week post-test, administered on the final day of the session.

(2) Introduction of a final evaluation sheet in lieu of the previous section evaluation sheets, due to the excess of unstructured, non-specific data provided by the previous forms which limited their usefulness for program revision and as indicators of student attitude. This sheet also collected specific unit-related and objective-related data to replace that eliminated by the deletion of the unit rating procedure.

(3) A controlled response terminal attitude measure was also added and collected three months following completion of training. It was felt that this procedure could provide more objective data, since students would have gained further experience and perspective toward the usefulness of their training.

CONCLUSIONS

As a result of the design and conduct of the formative evaluation within the particular educational context given and utilizing the particular evaluation model and procedures described, the following conclusions were drawn.

Formative evaluation must be "adaptive" evaluation. This implies that the evaluation design and procedures must be sufficiently flexible to adapt to the changing requirements of the particular evaluation context in which the evaluation is being conducted, and the evaluation model must allow for such flexibility. This is particularly significant in contexts such as the one described, where both initial program development and training of students are occurring simultaneously, where the students are being viewed as a primary source of input into the program development effort and where the evaluation functions in a service role requiring responsiveness to the operational requirements of the program.

The planning of the evaluation design and procedures must take into account any known characteristics of the particular students who comprise the pilot test group, in that they are a significant aspect of the evaluation context. Their evaluation preferences must be considered particularly in on-line revision of the evaluation procedures, in light of the importance of their willing cooperation in providing accurate and specific feedback into the evaluation of the instructional program for revisions purposes.

The CIPP model can serve as a useful base for overall program evaluation design, in that it provides a useful and inclusive conceptualization of and distinction among the formative evaluation processes to be conducted, a clear delineation of the evaluator role within the evaluation context, and an allowance for adaptation of the evaluation design and procedures during program implementation. It satisfied the evaluation staff's needs in these areas and all steps

proposed by the model for the process and product evaluation (formative only) could be conducted as implied by the model.

The timing of administration of the formative evaluation procedures can be extremely significant in terms of maximizing their reliability and effectiveness, especially when training is being conducted in concentrated segments. "Receptivity" to the evaluation procedures was higher, for example, at the beginning of the day or week than at the end, when students were fatigued and anticipating returning to their homes after a week's absence. One of the session posttests was eliminated due to staff realization that the conditions and timing duplicated those of an earlier posttest where results proved to be lower than on the pretest given at the beginning of the session. (Students attributed the lower scores to their fatigue after the intensity and duration of training.) The effectiveness of the unit rating sheets and daily logs was also a function of the timing of their administration and completion, e.g. the daily logs were initially to be collected at the end of each week. It was discovered, however, that students completed the forms only immediately prior to their due date resulting in the collection of very general data. The collection procedure was therefore altered to a daily schedule.

One danger in initial evaluation efforts such as the one undertaken here is that of evaluation "overkill," when too many evaluation procedures are assigned to collect similar kinds of information, and too many individuals are involved in its collection. This occurance had particularly negative effects upon the students since few of the procedures involved unobstrusive measures. The skeletal assignment of reliable procedures to gather the kinds of feedback desired is preferable both in terms of time required for the evaluation itself and in terms of analysis of the data received.

The division of the training program into three sessions separated by periods of time during which the students return to their jobs and attempt to apply the skills learned in the training sessions is an excellent structure for the conduct of a pilot test of a program, particularly where students are being

enjoined to assist in program development and validation of the program objectives. It enabled students to try out their new knowledges and skills and hence provide more accurate feedback about the usefulness of that training, and likewise allowed the program staff sufficient time to revise subsequent training components in light of the feedback received in previous sessions.

Certain evaluation procedures are more feasible and useful in terms of the provision of feedback than others. Feasibility was determined by whether a procedure could be successfully implemented as designed or conceived, a factor influenced by the degree of control exercised by the evaluation staff over the various program components combined with the availability of the necessary resources. Assessment of usefulness was viewed in terms of the type of feedback provided by the individual procedure, and how successfully and reliably this feedback was elicited. Exhibit 2 provides a rating of each procedure in terms of its effectiveness as implemented in the program for providing the kind of feedback specified, and a brief rationale for that rating in terms of the strengths or weaknesses of the particular procedure or instrument as designed and implemented. Exhibit 3 indicates those procedures which would be retained on future evaluations of the same program, with a brief description of any revisions that would be made.

In general, slightly more controlled response techniques, or a combination of controlled and free response techniques proved more useful and feasible in terms of amount of student time required, ease of data analysis, and usefulness in providing immediate feedback into program revision. Likewise, techniques such as the final evaluation sheet which requested specific, controlled feedback were more useful for revisions purposes than those soliciting controlled feedback on more general terms. Students also appeared to prefer them.

Techniques involving taping of feedback generally provided the problem of time-delay due to necessity of transcribing the data into print form before attempting to conduct a content analysis. This greatly hindered their usefulness.

EXHIBIT 2

Rating of Program 3A Evaluation Procedures for
Function Specified and Rationale

Immediate feedback		Rationale
High:	Unit ratings	Feedback differentiated, daily and representative
	Daily logs	Feedback differentiated, daily and representative
	Class feedback session	Feedback direct and based on perceived need; usually representative
Med:	Instructor observation	Potential bias, non-representative, non-systematic
	Evaluator observation	Potential bias, non-representative, non-systematic
	Individual feedback	Non-representative, and not in perspective
Low:	Taped interviews	Confidentiality of feedback, data access & analysis difficult

Delayed feedback for later revision		Rationale
High:	Self-assessment forms	Easily analyzed, objective-related, representative
	Unit ratings	Easily analyzed, objective-related, representative plus both free and controlled responses
	Objectives assessment	Easily analyzed, objective-specific, representative
	Final evaluation	Easily analyzed, representative, free and controlled responses, specific
Med to Low:	OJT Records	Unrepresentative, not completed properly
	Taped counselor interviews	Timing too early, too general, difficult to access and analyze data
	Class feedback sessions	Too general, uncontrolled, not unit-specific, access to taped data difficult

	Individual feedback	Unrepresentative, no systematic collection
	Daily logs	Too general, difficult to analyze, not always relevant
	Section evaluation	Too general, difficult to analyze, not always relevant

Achievement Assessment		Rationale
High:	Program pre/posttests	Direct relation to behavioral objectives
	Project evaluation	Direct relation to behavioral objectives
Med to Low:	Instructor observation	Assessed performance goals, but not systematically; potential bias
	OJT records	Unrepresentative, time-consuming
	Section 1 posttest	Timing of administration poor, so reliability low

Overall student attitude		
High:	Terminal attitude measure	Easily analyzed, representative, controlled
	Final evaluation	Easily analyzed, representative, controlled and free responding, specific
Med:	Taped counselor interview	Data access and analysis difficult
	Section evaluation	Too general, difficult to analyze not always relevant
	Daily logs	Too general, difficult to analyze, not always relevant, too time-consuming, redundant

Program monitoring		Rationale
Med:	Evaluator observation	Potential bias, non-representative, potentially unsystematic, time-consuming
Low:	Instructor rating	Not feasible due to time constraints
	Taping of instruction	Difficult to access & analyze data; but objective

EXHIBIT 3

Evaluation Procedures Retained and Type of Revision Planned

Immediate feedback

1. Unit ratings: more controlled, including objectives assessment and free attitudinal responses; more specific in terms of unit components, briefer.

2. Instructor or evaluator observation: depending upon type of instruction and if observation is relevant to assessment of particular objective type (affective, performance); more systematic, with forms provided.

3. Unit posttests: if objectives susceptible to formal direct assessment; very brief sampling; possible combining units into larger segments to avoid posttesting more often than once every two days.

4. Class feedback sessions: only if need indicated.

Delayed feedback for later revision

1. Unit ratings: more controlled, including objectives assessment and free attitudinal responses; more specific in terms of unit components, briefer.

2. Objectives assessment: of those from previous session at beginning of subsequent session following OJT and opportunity to assess value of objectives, or all of them at end of program.

3. OJT records: controlled, brief, and weekly, not daily; including project progress report.

4. Final evaluation: as given at end of 6 weeks.

5. Session posttests: of objectives susceptible to written assessment.

6. Interview: at end of program sessions, to gather specific information about student suggestions (on content, procedures, etc.) for subsequent sessions; systematically conducted, more controlled.

Achievement assessment

1. Program and session pretests: used diagnostically.

2. Unit and/or session posttests: preferable to one final posttest, due to length and interim OJT periods.

3. Project evaluation: more systematic.

4. Evaluator/instructor observation: of performance (non-product) objectives, during group or individual sessions; more systematically evaluated.

Overall student attitude

1. Terminal attitude measure: as given.

2. Final evaluation: as given at end of 6 weeks.

Program monitoring

1. Instructor rating sheet: to record any program changes as implemented.

2. Evaluator monitoring: only if needed and resources permit.

One important aspect of a formative evaluation effort which solicits student input is the maintenance of a positive attitude toward the evaluation effort and student participation in that effort. This is affected by such factors as degree of understanding of the purposes of the evaluation in which they are being asked to participate, the clarity of the explanation of the use of the individual procedures themselves, the degree to which the students are made to feel that they are participants in the program development process rather than merely program "guinea pigs," and the degree to which they feel that their feedback is being attended to by the developmental staff, as evidenced in either program modification based on that feedback, or direct feedback from the staff about its usefulness.

RECOMMENDATIONS

The problems encountered by the evaluation staff in its conduct of the training program evaluation have pinpointed several areas where further work in the area of formative educational evaluation would be useful.

A real need exists for the development of criteria for the pre and during-implementation evaluation of an evaluation plan and its component procedures. While effort in this direction has been initiated, as exemplified by Sorensen's "Formative evaluation checklist," (1971) and Stuffelbeam's 11 criteria (Stuffelbeam et. al., 1971), more work needs to be concentrated in this area.

There is a definite need for the development of validated formative evaluation instruments and evaluation guidelines which are context-specific, i.e. which assist in the selection and/or construction of procedures applicable to given program evaluation contexts. Although mention of the practical need to design evaluations in light of the evaluation context has been made (Alkin, 1971), the "individualization of evaluation" remains to be realized in terms of discovering meaningful procedural design/context interactions as well as in terms of proposing practical guidelines for such adaptation. These areas remain open to future exploration.

BIBLIOGRAPHY

Alkin, M. C. "Products for Improving Educational Evaluation," Evaluation Comment, 2 (3), 1970.

Popham, W. J. An Evaluation Guidebook: A Set of Practical Guidelines for the Educational Evaluator. The Instructional Objectives Exchange, Los Angeles, Calif.: 1971.

Sorensen, G. "Evaluation for the Improvement of Instructional Programs: Some Practical Steps." Evaluation Comment, 2 (4), 1971.

Stuffelbeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. L., Provus, M. M., Educational Evaluation and Decision Making. F. E. Peacock Publishers, Inc., Itaska, Illinois: 1971.

APPENDIX A

Description of Program 3A Evaluation Procedures

Monitoring Techniques

1. Evaluator observation: One member of the evaluation staff functioned in the role of program monitor, sitting in on all group instructional activities and keeping a record of whether the instruction followed the pre-planned activity flow and where modifications were made. A further extension of this role was the provision of more informally acquired feedback from individual students to the instructional staff.
2. Instructor rating: Forms were provided the instructor for the purpose of recording his assessment of the conduct of the instruction in each unit and to record any modifications made during implementation, data which would be useful for later program revision.
3. Taping of instructional sessions: All class sessions were audio taped on cassettes to assist in the monitoring of program implementation.

Student Opinion Techniques

1. Unit ratings: At the completion of each instructional unit (approximately two per day), the student was asked to fill out a form rating the instructional quality of the unit in terms of its various components, and to suggest possible revisions. Both controlled and free responses were solicited. Time was to be allotted at the end of each unit for this purpose, and the forms turned in immediately to the evaluator.

2. Daily logs: Each student was requested to record his general impressions of each day in anecdotal free-response form, guided only by such headings as "overall impressions," "problems encountered," etc. (Although these were originally to be collected at the end of each program section, it proved more practical to collect them each morning prior to commencing instruction.)

3. Section evaluation: At the end of each distinct section of the program (a total of four, each of which differed in instructional methodology and content orientation), students were requested to evaluate the section as an entity according to general guidelines such as "value of content to you as an LCS," or "problems encountered."

4. Counselor interviews (taped): As part of the guidance component, a feedback interview was conducted with each trainee during the second week of the program. Each student was invited to give his reactions to the program in line with guidelines established by the interviewer. The confidentiality of the interview was guaranteed to the student, with the exception of the project directors and evaluation staff.

5. Student self-assessment: Controlled-response rating sheets were provided to the students as part of two of the initial instructional units. Students were requested to rate their perception of their degree of knowledge in certain content areas or degree of attainment of certain specified skills related to the program goals. This procedure was used solely to gather background information about the trainees and their self-perceptions.

6. On-the-job training reports: During the intervening periods of work between sessions, students were requested to fill out brief daily and weekly reports on forms supplied. The daily report sheet was a mixture of free and controlled response items, whereas the

weekly report requested a summary of training-related activities (free response) based on daily records. These were to be submitted bi-monthly, and were to be responded to by the instructional staff. The trainee's immediate supervisor on the job was likewise asked to submit anecdotal bi-weekly reports on the trainee's progress.

7. Evaluation of unit objectives: During the fifth week of program implementation, students were asked to rate the value of the instructional objectives of the first 20 instructional units in terms of RG (generally relevant to LCS role), M (I mastered this objective), RJ (relevant to my job), and U (I used knowledge of objective on job.)
8. Class feedback sessions: Class sessions were on occasion thrown upon to permit general and informal discussion of the program content and didactics, both for purposes of gathering student reaction as well as to facilitate the creation of group (students and staff) unity and commitment to the program, its development and personal relevance. No structure was imposed on such sessions.
9. Individual feedback: Interactions with staff members frequently proved to be a vehicle for provision of informal student feedback on the conduct of the program and its objectives. No effort was made to collect these data systematically.

Achievement Assessment Techniques

1. Pre- and post-tests: Single-form achievement tests were developed to assess degree of student attainment of program (and occasionally section) objectives. Two pre-tests, one convering the entire program and a more specific one related to Section 1 only were administered at the onset of the six-week program. These were not used as placement or diagnostic instruments. A post-test followed completion of

the first section. No further formal testing occurred until the final week, when a review test similar to the initial pre-tests was administered. This was used to diagnose areas of student weaknesses to be corrected during the final week. A final brief post-test on the objectives covered in the final week plus any objectives failed in the review test was administered on the final day of the program.

2. Project evaluation: Each student was responsible for the design and conduct of an individual project. This was reviewed and evaluated by the instructors, and served to evaluate those objectives which were not amenable to test-item evaluation.
3. Instructor observation: Certain objectives related to affective performance required the use of more informal, unobstrusive measures. Such objectives were evaluated through instructor observation during role-playing activities, class discussion, and during formal tutorials and interviews with the instructors.

Student Attitude Measures

1. Final evaluation sheet: A final evaluation sheet was distributed at the end of the 6th week of the program. A mixture of controlled and free-response questions solicited specific information about trainee reaction as to whether time spent in the program was worth the time absented from work, whether any of the training materials had been useful in the individual's work, etc.
2. Terminal attitude measure: Three months after program completion, students were sent a 10-item rating sheet requesting the rating on a 5-point scale of such questions as (1) whether the program was worth the six weeks spent on it, (2) whether the necessary materials were available to meet the stated objectives, etc.